

CHALLENGES IN AUTOMATION OF QUALITY CONTROL FOR TIDE GAUGE DATA

Felix Soltau, University of Siegen, Germany, felix.soltau@uni-siegen.de
Sebastian Niehüser, University of Rostock, Germany, sebastian.niehueser@uni-rostock.de
Jürgen Jensen, University of Siegen, Germany, juergen.jensen@uni-siegen.de

INTRODUCTION

Tide gauges provide important water level data for navigation, port management, coastal protection strategies, ecological adaptation measures, or climate change assessments. For these tasks, a reliable availability and high quality of the data is crucial. However, water level data from tide gauges contain technical errors as well as anthropogenic and natural influences. For the German North Sea coast and estuaries, resulting water level anomalies are partially detected and corrected manually by qualified personnel and further considered by individual subsequent users of that data. Figure 1 shows an example of such a correction of water level anomalies around tidal low water from tide gauge data at Husum, Germany, in 2016. In general, manual quality control leads to different handlings and thus incomparable results. Consequently, a uniform and automated pre-processing is needed for tide gauge data in Germany in order to detect, correct, and classify anomalies ideally in real time. The developed pre-processing approaches will not be limited to tide gauges in Germany but can be globally transferred or be extended to river sites.

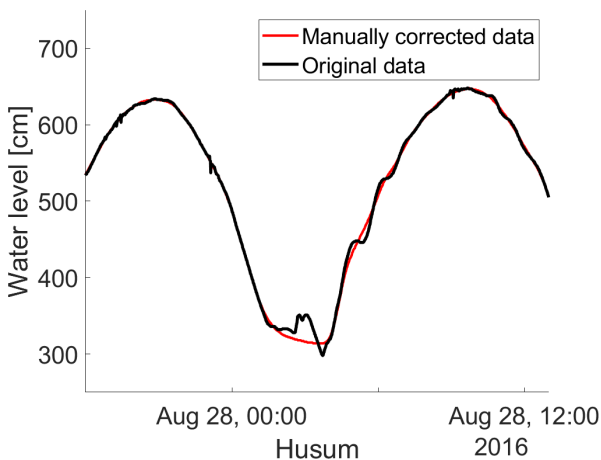


Figure 1 - Example of manual correction of water level anomalies around tidal low water from tide gauge data at Husum, Germany, in 2016

METHODOLOGY

In the research project DePaRT*, we focus on automated detection, classification, and correction of water level anomalies in tide gauge data located at the German North Sea coast and the estuaries. For anomaly detection, we generate and adapt a variety of statistical methods that check the measured water levels and recognize technical malfunctions of the gauge resulting in outliers or gaps, effects from anthropogenic influences like operating storm surge barriers and sluices, or effects from natural phenomena like standing waves, meteotsunamis, or inland flood events. The methods comprise filters validating single

values, multiple-value evaluation, or statistical models on longer sequences of water level data. Single-value filter detect non-physical measurements like negative water levels. Multiple-value evaluations validate a value by including adjacent values and measuring the difference from an expected water level development. Statistical models like artificial neural networks detect effects of anthropogenic and natural influences on water level data. They constitute a substantial part in this study. We use sequences of water level time series containing both anomalies and ordinary water levels to train them while using different architectures of networks. As broad training data sets are needed for these non-linear statistical methods, we synthesize rarely observed anomalies to gain a large number of artificial representatives. All methods are used consecutively and detect all kinds of obvious errors to possible anomalies due to natural effects.

For classification and quality assessment of the water level anomalies as well as for corrected values, we introduce classes. The values of the water level time series are assigned to the classes depending on the different methods. A documentation of the methods and related assignments prevents individual interpretations of water levels and ensures a uniform handling of values and transparency of possible modifications from correction methods for subsequent users.

CONCLUSION

Tide gauge data at the German North Sea coast are presently validated and corrected manually, which is time consuming, error-prone and cost-intensive. Here, we present a combination of consecutive statistical methods to introduce automation of quality control of tide gauge data. Parallel to this study, we implement the developed algorithms in an operational test mode at the German authorities' systems, while cooperating closely with the personnel to benefit from the mutual feedback. Hence, the resulting uniform and automated pre-processing for tide gauge data has a high practical relevance. It contributes to significantly enhance data quality. This major advance also reduces the uncertainties in the subsequent analysis and hence increases the robustness of the results.

* DePaRT (Detektion von küstenhydrologischen Phänomenen und Artefakten in minütlichen Tidepegeldaten; engl. Detection of coastal hydrological phenomena and artefacts in minute-by-minute tide gauge data) is a research project, funded by the German Federal Ministry of Education and Research (BMBF) through the project management of Projektträger Jülich PTJ under the grant number 03KIS133.